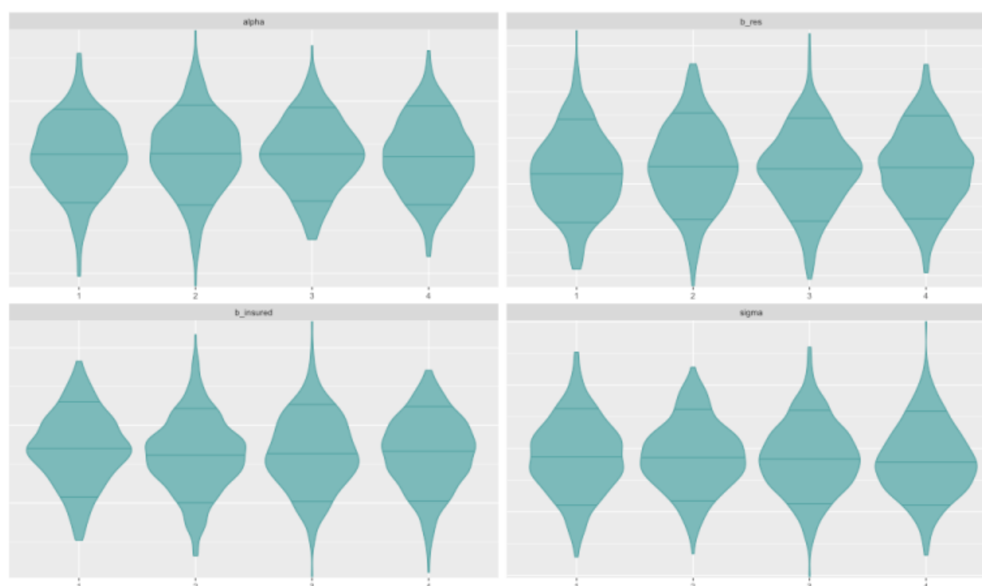
Chương trình máy tính bayesvl trong môi trường R: Đóng góp Việt cho khoa học thế giới

Hoàng Phương Hạnh (Viện Khoa học Giáo dục Việt Nam) *13/06/2019 13:35*

https://khoahocphattrien.vn/khoa-hoc/chuong-trinh-may-tinh-bayesvl-trong-moi-truong-r-dong-gop-viet-cho-khoa-hoc-the-gioi/20190613093044113p1c160.htm

Mới đây, chương trình máy tính 'bayesvl' chạy trên môi trường R do TS. Vương Quân Hoàng và kĩ sư Lã Việt Phương (Trung tâm Nghiên cứu Xã hội Liên ngành ISR, Đại học Phenikaa) thiết kế và phát triển chính thức được ra mắt trên CRAN - hệ thống thư viện chuẩn của R - tại địa chỉ xuất bản: https://cran.r-project.org/package=bayesvl.



Hình ảnh được sản xuất bởi chương trình bayesvl.

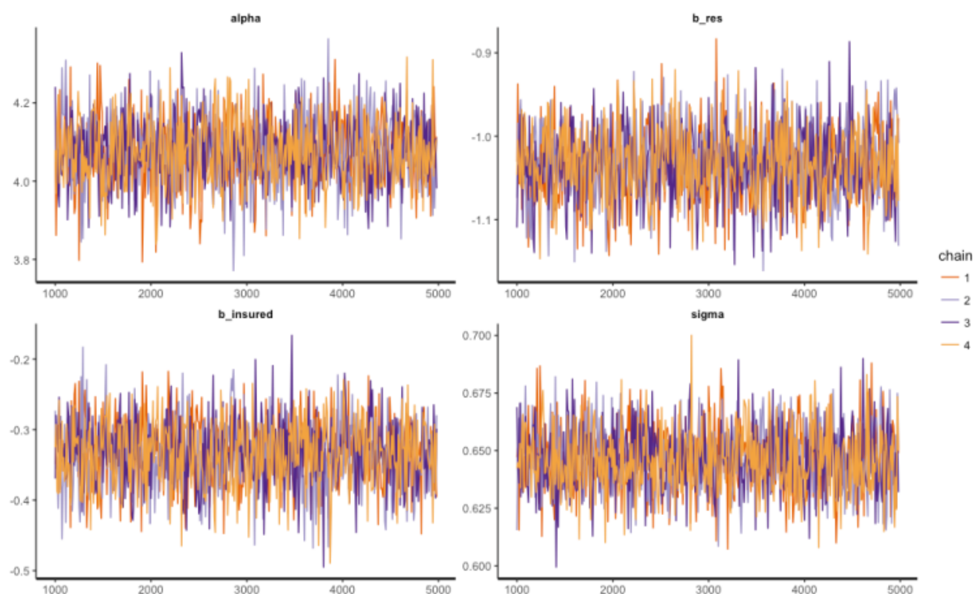
Chương trình bayesvl được thiết kế với định hướng sư phạm, hỗ trợ người sử dụng làm quen với phương pháp thống kê Bayesian dựa trên mô hình lưới Bayesian, mô phỏng MCMC, hình ảnh hóa các thông số kĩ thuật và kết quả.

Có thể hiểu đơn giản bayesvl giúp dựng mô hình rồi tạo ra code stan, và sau đó tận dụng các hàm vẽ đồ họa của chương trình. Hướng tới khoa học mở, ngay từ trong quá

trình phát triển, chương trình đã được các tác giả đăng tải trên trang chia sẻ phần mềm và code GitHub.

Kĩ sư Lã Việt Phương, một trong hai tác giả phát triển chương trình, cho biết: "bayesvl được lên ý tưởng từ cuối năm 2017 đến nay. Trong thời gian đó, phương pháp Bayesian đã được Trung tâm ISR nghiên cứu, thử nghiệm, và sử dụng thành công trong các công bố khoa học của trung tâm. Tuy nhiên, việc ứng dụng phương pháp này cần nhiều thời gian làm quen do sự phức tạp của các chương trình đã được phát triển sẵn như rstan.

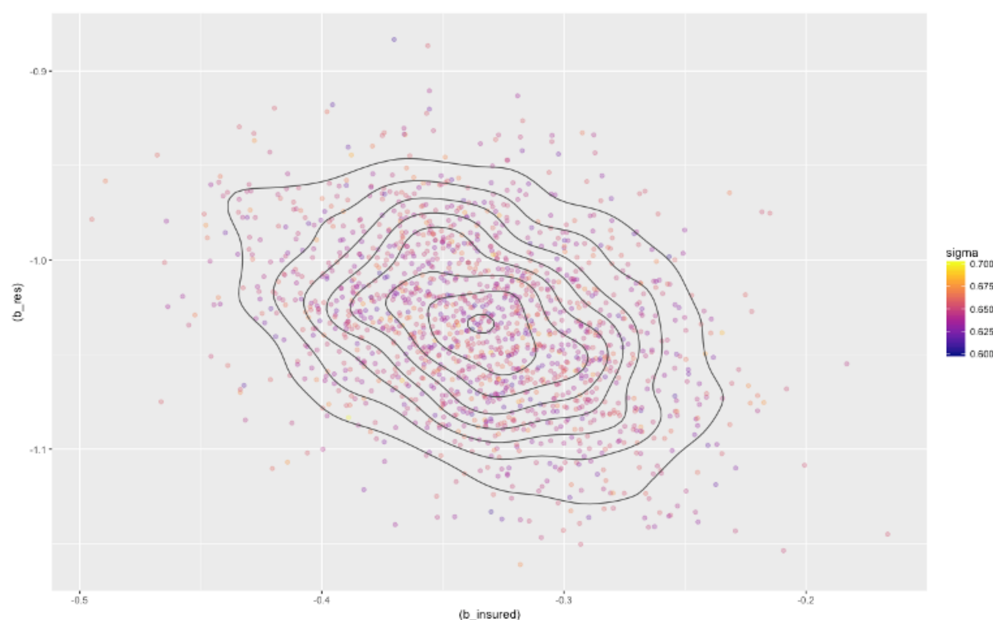
Chúng tôi nhận thấy nhu cầu thiết thực cho một chương trình thống kê Bayesian dễ sử dụng, nên đã bắt tay vào thiết kế và triển khai dự án. Sau hơn 18 tháng, và hơn 3000 dòng code, chương trình bayesvl cuối cùng đã chính thức được CRAN chấp nhận và xuất bản."



Hình ảnh được sản xuất bởi chương trình bayesvl.

Đầu năm 2018 đến nay, Trung tâm ISR đã áp dụng phương pháp Bayesian thành công trong nghiên cứu về truyện cổ tích và văn hóa Việt Nam trên tạp chí Palgrave Communications thuộc Nature Research. Sau đó, từ đầu năm 2019 đến nay, trung tâm đã sử dụng bayesvl trong 2 công bố trên tạp chí Data [ESCI; Scopus] thuộc nhà xuất bản MDPI, và một bản thảo về tiếp biến văn hóa và phố cổ Hà Nội đang trong quá trình phản biện.

Ngày 18/5/2019, chương trình được nộp vào hệ thống CRAN và sau 4 lần chỉnh sửa theo yêu cầu của kĩ thuật viên của CRAN, phiên bản 0.8.5 chính thức được xuất bản vào ngày 24/5/2019 với tên gọi: bayesvl: Visually Learning the Graphical Structure of Bayesian Networks and Performing MCMC with 'Stan'.



Hình ảnh được sản xuất bởi chương trình bayesvl.

Anh Nguyễn Minh Hoàng, du học sinh tại trường Đại học Ritsumeikan Asia Pacific University, Nhận Bản, đã sử dụng bayesvl từ phiên bản 0.6 khi chương trình vẫn còn trong giai đoạn phát triển. Anh nhận xét: "Mình đã tìm đọc và làm quen với phương pháp Bayesian trong suốt thời gian nghiên cứu thạc sĩ từ năm 2017. Tuy nhiên, việc ứng dụng nó vào thực tế lại rất khó khăn do bản thân phương pháp, dù mạnh mẽ, nhưng chưa phổ biến do sự phức tạp của các yếu tố toán học và code máy tính. Chính vì vậy, ngay khi biết đến bayesvl trên GitHub, mình đã thử sử dụng theo hướng dẫn.

Chương trình này được thiết kế dễ hiểu, giảm thiểu thời gian làm quen tiếp cận để sử dụng thành thạo code máy tính. Đồng thời, bayesvl cũng kích thích người dùng đầu tư vào công đoạn thiết kế đầu bài cũng như mô hình tính toán để quá trình triển khai thuận lợi. Mình khá ưng ý với sư đầu tư vào sản xuất hình ảnh của chương trình. Mô hình hay kết quả đều được thể hiện rõ ràng qua đồ họa, giúp quá trình làm luận án của mình thuận lợi hơn rất nhiều."

Trong thời đại công nghệ 4.0 hiện nay, ngay cả các nghiên cứu lĩnh vực KHXH&NV cũng đang dần chuyển mình trước làn sóng dữ liệu, khoa học mở, và phương pháp thống kê. Khi các tranh luận về ý nghĩa thống kê và tái xác lập kết quả vẫn đang là chủ đề rất được quan tâm đối với giới khoa học toàn cầu, chương trình bayesvl do hai tác giả Việt Nam xuất bản là một đóng góp quan trọng về phương pháp luận cho cộng đồng khoa học Việt Nam và thế giới. Trong tương lai, các tác giả hy vọng sẽ có thêm nhiều nhà khoa học trong và ngoài nước tìm đến và sử dụng bayesvl.

R là ngôn ngữ lập trình và môi trường phần mềm hoàn toàn miễn phí được thiết kế cho tính toán thống kê và đồ họa. Hiện nay, R được sử dụng rộng rãi cho việc phát triển các chương trình thống kê và phân tích dữ liệu. Là ngôn ngữ lập trình mã nguồn mở, các tính năng cho R luôn được mở rộng và cải thiện thông qua các gói phần mềm (package), thường là do người dùng thiết kế và xây dựng. Các gói phần mềm này mở rộng thêm tính năng thống kê hay sản xuất hình ảnh, phục vụ nhu cầu phức tạp và đa dạng của người dùng.

CRAN hay Comprehensive R Archive Network là thư viện lưu trữ, phân phối các phiên bản chính thức của R và hơn 15.000 gói phần mềm, hay các phần code bổ sung cho R. Bên cạnh CRAN, người dùng còn có thể tìm kiếm thêm các gói phần mềm này trên các hệ thống khác như GitHub, Omegahat hay Bioconductor.

Hoàng Phương Hạnh (Viện Khoa học Giáo dục Việt Nam)